METHOD FOR OBTAINING DNA, RNA, PEPTIDES, POLYPEPTIDES OR PROTEINS BY
MEANS OF A DNA RECOMBINANT TECHNIQUE

Marc Ballivet, Stuart Kauffman

INTERNATIONAL PATENT OFFICE

WORLD ORGANIZATION FOR INTELLECTUAL PROPERTY

International patent published on

the basis of the Patent Cooperation Treaty

INTERNATIONAL PUBLICATION NO. WO 86/05803 A1

METHOD FOR OBTAINING DNA, RNA, PEPTIDES, POLYPEPTIDES OR PROTEINS BY MEANS OF A DNA RECOMBINANT TECHNIQUE

[Procede d'obtention d'adn, arn, peptides, polypeptides ou proteines, par une technique de recombinaison d'adn]

Inventor and
Inventor/Applicant: Marc Ballivet and
                                    Stuart Kauffman

Designated States: AT (European patent), AU, BE
                                    (European patent), CF (OAPI
                                    patent), CG (OAPI patent), CH
                                    (European patent), CM (OAPI
                                    patent), DE, DE (auxiliary utility
                                    model), DE (European patent), FR
                                    (European patent), GA (OAPI

patent), GB, GB (European patent),
IT (European patent), JP, LU
(European patent), ML (OAPI
patent), MR (OAPI patent), NL
(European patent), SE (European
patent), SN (OAPI patent), TD
(OAPI patent), TG (OAPI patent),
US.

Published
    With International Search Report

FOR INFORMATION ONLY
Codes for the identification of PCT contract states on the cover sheets of the
documents that publish the international applications in accordance with the PCT.

AT	Austria	US	United States of
AU	Australia		America
BB	Barbados		
BE	Belgium		
BF	Burkina Faso		
BG	Bulgaria		
BR	Brazil		
CF	Central African Republic		
CG	Congo		
CH	Switzerland		
CM	Cameroon		
DE	Germany		
DK	Denmark		
FI	Finland		
FR	France		
GA	Gabon		
GB	United Kingdom		
HU	Hungary		
IT	Italy		
JP	Japan		
KP	Democratic People's Republic of Korea		
KR	Republic of Korea		
LI	Liechtenstein		
LK	Sri Lanka		
LU	Luxembourg		
MC	Monaco		
MG	Madagascar		
ML	Mali		
MR	Mauritania		
MW	Malawi		
NL	Netherlands		
NO	Norway		
RO	Romania		
SD	Sudan		
SE	Sweden		
SN	Senegal		
SU	Soviet Union		
TD	Chad		
TG	Togo		

The goal of the present invention is a method for obtaining DNA, RNA, peptides, /1[*] polypeptides or proteins by means of modified host cells containing genes capable of expressing these RNAs, peptides, polypeptides or proteins, that is, by using a DNA recombination technique.

The invention aims especially to produce stochastic genes or fragments of genes to enable, after transcription and translation of these genes, simultaneously obtaining a very large number (on the order of at least ten thousand) completely new proteins or hybrids of known proteins, in the presence of host cells (bacterial strains or eukaryotes) containing the genes respectively capable of expressing these proteins and of then carrying out a selection or screening among said strains, with a view to determining those that produce proteins presenting the desired properties, for example structural enzyme, catalytic, antigenic, pharmacological properties or ligand properties, and more generally chemical, biochemical, biological etc. properties.

The invention also has the purpose of enabling sequences of DNA or RNA to be obtained that have usable properties, especially chemical, biochemical or biological properties.

Therefore, it is understood that the invention is likely to find very numerous applications in very varied fields of science, industry and medicine.

The production method of peptides or polypeptides according to the invention is /2 characterized in that one produces simultaneously, within the same medium, genes that are at least partially composed of stochastic synthetic polynucleotides, that the genes so obtained are introduced into host cells, that the independent strains of modified host cells containing these genes are cultured simultaneously so as to clone the stochastic genes and to promote the production of genes expressed by each of the stochastic genes, that screening and/or selection of the strains of modified host cells is carried out so as to identify the strains producing peptides or polypeptides with at least one given property, that the strains so identified are isolated and that they are cultured so as to produce at least one peptide or polypeptide with said property.

In conformance with a first embodiment of the method, the genes are produced by stochastic copolymerization from four types of deoxyphosphonucleotides A, C, G and T from two ends of a previously linearized expression vector, then formation of cohesive ends so as to form a first strand of stochastic DNA formed from a molecule of the expression vector containing two stochastic sequences, the 3' ends of which are complementary, followed by the synthesis of the second strand of this stochastic DNA.

---

[*] [The number in the margin indicates the pagination of the foreign text.]

According to a second embodiment, the genes are produced by stochastic copolymerization of oligonucleotides without cohesive ends, so as to form fragments of stochastic DNA, followed by ligation of these fragments to a previously linearized expression vector.

The expression vector may be a plasmid, especially a bacterial plasmid. Excellent results have been obtained by using the plasmid pUC8 as expression vector.

The expression vector may also be a viral DNA or a hybrid of plasmid and viral DNA. /3

The host cells may be prokaryotic cells such as the HB 101 and C 600 cells or eurkaryotic cells.

When the method is conducted in conformance with the second embodiment mentioned above, oligonucleotides may be used forming an assembly of palindromic octamers.

Particularly good results are obtained by using the following group of palindromic octamers:

```
5' GGAATTCC 3'
5' GGTCGACC 3'
5' CAAGCTTG 3'
5' CCATATGG 3'
5' CATCGATG 3'
```

Oligonucleotides may also be used that form an assembly of palindromic heptamers.
Very good results are obtained by using the following group of palindromic heptamers:

```
5' XTCGCGA 3'
5' XCTGCAG 3'
5' RGGTACC 3'
```

where x = A, G, C or T and R = A or T.

In conformance with a particularly advantageous embodiment, the transformant DNA of the plasmids that originates from a culture of independent strains of modified host cells obtained by proceeding in the manner specified above is isolated and purified, then the cleavage of the DNA is promoted by means of at least one restriction enzyme corresponding to a specific enzymatic cleavage site present in these octamers or heptamers but absent from the expression /4

vector used, this cleavage being followed by the inactivation of the restriction enzyme, and then all the fragments of linearized stochastic DNA so obtained are treated simultaneously with T4 DNA ligase to create a new DNA assembly that contains new stochastic sequences, this new assembly therefore being able to contain a number of stochastic genes greater than the number of genes of the initial assembly, and this new assembly of transformant DNA is used to modify host cells and clone genes, and finally, the new strains of transformed host cells are screened and/or selected and isolated and finally they are cultured to produce at least one peptide or polypeptide, for example a new protein.

The properties acting as selection criterion of the strains of host cells may be the ability of the peptides or polypeptides produced by this strain to catalyze a given chemical reaction.

For example, for the production of several peptides and/or polypeptides said property may be the ability to catalyze a sequence of reactions leading from a given initial group of chemical compounds to at least one target compound.

Considering the production of an assembly formed from several reflexively autocatalytic peptides and/or polypeptides, said property may be the ability to catalyze the synthesis of this assembly itself from amino acids and/or oligopeptides in an appropriate medium.

Said property may also be the ability to selectively modify the chemical and/or biological properties of a given compound, for example the ability to selectively modify the catalytic activity of a polypeptide.

Said property may also be the ability to simulate, inhibit or modify at least one biological function of at least one biologically active compound chosen for example from hormones, neurotransmitters, adhesion or growth factors and the specific regulators of DNA replication and/or transcription and/or translation of RNA.

Said property may also be the ability of the peptide or polypeptide to be bound to a given ligand.

The invention also has as a goal the use of the peptide or polypeptide obtained by the method specified above for the detection and/or titration of a ligand.

In conformance with one particularly advantageous embodiment, the selection criterion of the modified strain of host cells is the ability of the peptides or polypeptides to simulate or modify the effects of a biologically active molecule, for example a protein, and the screening and/or selection of the modified strain of host cells producing at least one peptide or polypeptide having this property by preparing antibodies against this active molecule, and by using these antibodies after their purification to identify the strains containing this peptide or polypeptide, then by culturing the strains so identified and by separating and purifying the peptide or polypeptide produced by these strains and finally, by subjecting this peptide or polypeptide to an

in vitro test to verify that it really has the ability to simulate or modify the effects of said molecule.

In conformance with another embodiment of the method according to the invention, the property acting as selection criterion is from having at least one epitope similar to one of the epitopes of a given antigen.

The invention also relates to the polypeptides obtained by the method specified above, and usable as chemotherapy active substance.

In particular, in the case where said antigen is EGF, the invention makes it possible to obtain polypeptides that are usable for the chemotherapy treatment of epitheliomas.

According to a variant of the method, the modified strains of host cells producing the peptides or polypeptides with the desired property are identified and isolated by affinity chromatography on antibodies corresponding to a protein expressed by the natural part of the hybrid DNA.

For example, in the case where the natural part of the hybrid DNA contains a gene expressing β-galactosidase, said modified strains of host cells may be advantageously identified by affinity chromatography on anti-β-galactosidase antibodies.

After expression and purification of the hybrid peptides or polypeptides, their new parts may be separated and isolated.

The invention also relates to an application of the method specified above for the preparation of a vaccine, this application being characterized by the fact that antibodies against a pathogenic agent are isolated, for example antibodies formed after injection of this pathogenic agent into the body of an animal capable of forming antibodies against this agent, and these antibodies are used to identify the clones producing at least one protein having at least an epitope similar to one of the epitopes of the pathogenic agent, the strains of modified host cells corresponding to these clones are cultured, in order to produce this protein, this protein is isolated and purified from cultures of these strains of cells, and this protein is used for the production of a vaccine against the pathogenic agent.

For example, for the preparation of an anti-HBV vaccine, at least one HBV virus capsid protein can be extracted and purified and this protein injected into the body of an animal capable of forming antibodies against this protein, the antibodies so formed are collected and purified, these antibodies are used to identify the clones producing at least one protein with at least one epitope similar to one of the epitopes of the HBV virus, the strains of modified host cells corresponding to these clones are cultured, to produce this protein, this protein is isolated and purified from cultures of these strains of cells, and this protein used for production of an anti-HBV vaccine.

Advantageously, in conformance with one embodiment of the method according to the invention, the host cells consist of bacteria of the genus Escherichia coli, the genome of which contains neither the natural gene expressing β-galactosidase nor the EBG gene, that is, (Z⁻, EBG⁻) E. coli bacteria, the modified host cells are cultured in the presence of the X-gal medium and the IPTG inducer, the clones positive for the β-galactosidase function are detected in the culture medium and finally, this DNA is transplanted into an appropriate strain of host cells for culture in large quantity, with a view to the industrial production of at least one peptide or polypeptide.

The property acting as selection criterion for the strains of transformed host cells may also be the ability of the polypeptides or peptides that are produced by culture of these strains to be bound to a given compound.

This compound may especially be chosen from peptides, polypeptides and proteins, especially the regulatory proteins of DNA transcription activity.

/8

On the other hand, said compound may also be chosen from the DNA and RNA sequences.

The invention also has as a goal the proteins obtained in the case where the property acting as selection criterion of the strains of transformed host cells consists precisely in the ability of these proteins to be bound to regulatory proteins of DNA transcription activity, or even DNA or RNA sequences.

In addition, the invention has as a goal the use of a protein obtained in the first specific case which has just been mentioned, as cis-regulatory of replication or transcription of a close DNA sequence.

On the other hand, the invention also has as a goal the use of proteins obtained in the second specific above-mentioned case to modify the transcription or replication properties of a sequence of DNA in a cell containing this DNA sequence and expressing this protein.

The invention also has as a goal a method of production of DNA, characterized by the fact that simultaneously and within the same medium genes are produced that are at least partially composed of stochastic synthetic polynucleotides, that the genes so obtained are introduced into host cells to produce an assembly of modified host cells, that screening and/or selection of this group is carried out in order to identify the host cells that contain the desired sequences in their genome of stochastic sequences of DNA presenting at least one desired property and finally, the DNA is isolated from cultures of host cells so identified.

/9

In addition, the invention has as a goal a production method of RNA characterized by the fact that simultaneously and within the same medium, genes are produced that are at least partially composed of stochastic synthetic polynucleotides, that the genes so obtained are introduced into host cells to produce an assembly of modified host cells, that the independent

strains of modified host cells so produced are simultaneously cultured, that screening and/or selection of this group is carried out in order to identify the host cells that contain stochastic sequences of RNA with at least one desired property, and that the RNA is isolated from cultures of host cells so identified.

Said properties may advantageously be the ability to be bound to a given compound, which may be for example, a peptide, polypeptide or a protein or even the ability to catalyze a given chemical reaction or even the property of being a transfer RNA.

The method according to the invention will now be described in more detail as well as some of its applications, by referring to nonlimiting application examples.

First, particularly advantageous operating methods will be described to carry out the synthesis of stochastic genes and the introduction of these genes into bacteria, so as to produce strains of transformed bacteria.

I Direct synthesis on an expression vector

a) Linearization of the vector

30 µg (that is, approximately 10^{13} molecules) of the expression vector pUC8 are   /10 linearized by incubation for 2 h at 37 ° (with 100 units of restriction enzyme PstI in a volume of 300 µL of appropriate standard buffer. The linearized vector is treated with phenol-chloroform then precipitated with ethanol, taken up in a volume of 30 µL and loaded on a 0.8% agarose gel in standard TEB medium. After migration in a field of 30 V/cm for three hours, the linear vector is electroeluted, precipitated in ethanol and taken up in 30 µL of water.

b) Stochastic synthesis by the enzyme, Terminal-Transferase (TdT)

30 µg of linearized vector are reacted with 30 units of TdT in 300 µL of appropriate buffer for two hours at 37 °C, in the presence of 1 mM of dGTP, 1 mM of dCTP, 0.3 mM of dTTP and 1 mM of dATP: a lower concentration of dTTP is chosen for the purpose of decreasing the frequency of "stop" codons in the corresponding messenger RNA. A similar result, although less favorable, may be obtained by using a lower concentration for the ATP than for the other deoxynucleotide triphosphates. The progress of the polymerization reaction on the 3' end of the Pst I sites is followed by gel analysis of aliquot parts sampled during the reaction.

When the reaction reaches or exceeds an average value of 300 nucleotides per 3' end, it is interrupted and the free nucleotides are separated from the polymer by differential precipitation or by passage on a molecular sieve of the Biogel P60 type. After concentration by precipitation in ethanol, the polymer is sequentially subjected to an additional polymerization by TdT first in the presence of dATP, then dTTP. These last two reactions are separated by a gel filtration and   /11

conducted over short times (30 sec to 3 min) to lead to the sequential addition of 10-30 A, followed by 10-30 T at the 3' end of the polymers.

c) Synthesis of the second strand of the stochastic DNA

At the end of the previous operations each vector molecule has two stochastic sequences, the 3' ends of which are complementary. The mixture of polymers is then incubated under conditions favoring hybridization of these complementary ends (150 mM NaCl, 10 mM Tris-HCl, pH 7.6, 1 mM EDTA, at 65 °C for 10 min, then lowering the temperature to 22 °C, at the rate of 3-4 °C /h). The hybridized polymers are then subjected to the action of 60 U of the large fragment of polymerase I (Klenow) in the presence of the four nucleotide triphosphates (200 mM) at 4 °C, for 2 h. This step carries out the synthesis of the second strand from the 3' end of the hybridized polymers. The molecules resulting from this direct synthesis from a linearized vector are then used to transform competent cells.

d) Transformation of the competent strains

100-200 mL of competent cells HB101 or C600 at the concentration of 10^{10} cells/mL are incubated with the preparation of stochastic DNA in the presence of 6 mM $CaCl_2$, 6 mM Tris-HCl, pH 8, 6 mM $MgCl_2$, for 30 min at 0 °C. Thermal shock for 3 min at 37 °C is applied to the mixture, followed by addition of 400-800 mL of NZY culture medium without antibiotic. The transformed culture is incubated at 37 °C for 60 min, and then diluted to 10 L by addition of the NZY medium containing 40 µg/mL of ampicillin. After 3-5 h of incubation at 37 °C, the amplified culture is subjected to centrifugation and the pellet of transformed cells is lyophilized and stored at -70 °C. Such a culture consists of 3 x 10^7 to 10^8 independent transformants, each /12 containing a single stochastic gene inserted in an expression vector.

II) Synthesis of stochastic genes from oligonucleotides without cohesive ends

This operating method is based on the fact that the polymerization of carefully chosen palindromic oligonucleotides allows assembly of stochastic genes not including any "stop" codon in each of the six possible reading frames, while ensuring equilibrated representation of triplets representing all the amino acids. Moreover, and to avoid repetition of the sequence units in the resulting protein, the oligonucleotides can include a number of bases that are not a multiple of three. The following example describes the use of a possible combination of oligonucleotides that fulfills these criteria:

a) Choice of an assembly of octamers

The following group of oligonucleotides:

```
5' GGAATTCC 3'
5' GGTCGACC 3'
5' CAAGCTTG 3'
5' CCATATGG 3'
5' CATCGATG 3'
```

is composed of 5 palindromes (including autocomplementary) for which it is easy to verify that their stochastic polymerization does not create a "stop" codon and specifies all the amino acids.

Of course, other assemblies of palindromic octamers could also be used that do not create "stop" codons and that specify all the amino acids of the polypeptides. Also, it will be understood that nonpalidromic assemblies of octamers could be used on condition that their complements also are used by the formation of bicatenary segments.

/13

b) Assembly of a stochastic gene from an assembly of octamers

A mixture including 5 μg of each of the oligonucleotides given above (previously phosphorylated at 5' by a method known in itself) in a volume of 100 μL containing 1 mM ATP, 10% polyethylene glycol and 100 u of T₄DNA ligase in the appropriate buffer at 13 °C, for 6 h. This step carries out the stochastic polymerization of the oligomers in bicatenary form, with cohesive ends. Then the polymers are isolated by passing the polymers resulting from the assembly of 20-100 oligomers through a molecular sieve (Biogel P60). After concentration, this fraction is again subjected to polymerization catalyzed by T4 DNA ligase under the conditions described above. Then, as described above, the polymers resulting from the assembly of at least 100 oligomers are isolated.

c) Preparation of the host plasmid

The expression vector pUC8 is linearized by the enzyme SmaI in the appropriate buffer as described above. The vector linearized by the enzyme SmaI does not include cohesive ends. the linearized vector is then treated with alkaline phosphatase from veal intestine (CIP) at one unit per μg of vector in the appropriate buffer at 37 °C for 30 min. The CIP enzyme is then inactivated by two successive extractions with phenol-chloroform. The linearized and dephosphorylated vector is precipitated in ethanol then taken up in water at 1 mg/mL.

d) Ligation of the stochastic genes to the vector

Equimolar quantities of vector and polymer are mixed and incubated in the presence of /14
1000 u of T4 DNA ligase, 1 mM ATP, 10% polyethylene glycol in the appropriate buffer for
12 h at 13 °C. This step carries out the ligation of the polymers in the expression vector and
creates bicatenary circular molecules, and as a result, transformants.

Transformation of the competent strains

The transformation of the competent strains is carried out in the previously described
way.

III Assembly of the stochastic gene from an assembly of heptamers

This operating method is distinguished from that which has just been described by the
fact that it uses palindromic heptamers including a variable cohesive end, instead of octamers. It
has the advantage of enabling the assembly of stochastic sequences including a lower proportion
of identical units.

a) Choice of an assembly of heptamers

For example, the following 3 palindromic heptamers may be used:

$$5' \quad XTCGCGA \quad 3'$$
$$5' \quad XCTGCAG \quad 3'$$
$$5' \quad RGGTACC \quad 3'$$

for which X = A, G, C or T and R = A, or T and in which the polymerization cannot create
"stop" codons and includes triplets specifying all the amino acids.

Of course, any other group of heptamers may be used that fulfill these same conditions.

b) Polymerization of an assembly of heptamers /15

This polymerization is carried out in exactly the way described previously for the
octamers.

c) Removal of the cohesive ends

The polymers so obtained include an unpaired base at their 5' ends. Therefore, it is
necessary to add the complementary base at the corresponding 3' ends. This is carried out in the
following way: 10 μg of double-strand polymers are reacted with 10 u of Klenow enzymes in the

presence of four deoxynucleotide phosphates (200 mM) in a volume of 100 µL at 4 °C for 60 min. The enzyme is inactivated by extraction with phenol-chloroform and the polymers are stripped of the residual free nucleotides by differential precipitation. The polymers are then ligated to the host plasmid (previously linearized and dephosphorylated) by proceeding in the manner described above.

Note that both last operating methods that have just been described use palindromic octamers or heptamers that form specific restriction enzyme sites. These sites are absent for the most part from expression vector pUC8. Therefore, it is possible to considerably increase the complexity of the initial preparation of stochastic genes by proceeding in the following way: the DNA is prepared from plasmids originating from a culture of 10^7 independent transformants obtained by one of the last two operating methods described above. This purified DNA is then subjected to partial digestion by the restriction enzyme Cla I (operating method II) or by the restriction enzyme Pst I (operating method III). After inactivation of the enzyme, the partially digested DNA is treated with T4 DNA ligase, which has the effect of creating an extremely large number of new sequences that keep the fundamental properties of initial sequences. This new   /16
assembly of stochastic sequences is then used for transforming competent cells.

Moreover, the cloned stochastic genes by proceeding in conformance with operating methods II and III may be excised intact from the expression vector pUC8 by using the appropriate restriction sites for the cloning vector and not represented in the stochastic DNA.

The recombination within the stochastic genes created by the last two operating methods that have just been described, recombination that results from the internal homology and recurrence units, provides an important additional method of mutagenesis in vivo of coding sequences. From this, there results an increase in the number of new genes that can be examined.

Finally, for any process for creation of new synthetic genes, numerous usual techniques for modification of genes in vivo or in vitro may be used, such as changing the reading frame, inversion of sequences relative to their promoter, or use of host cells expressing one or more suppressor tRNAs.

Referring to the preceding part of the description, it will be understood that it is possible to construct in vitro an extremely large number (for example, greater than one billion) different genes, by enzymatic polymerization of nucleotides and oligonucleotides. This polymerization is carried out according to a stochastic procedure determined by the respective concentrations of nucleotides or oligonucleotides present in the reaction medium.

As indicated above, two methods may be used to clone such genes (or coding sequences): the polymerization may be carried out directly on an expression cloning vector, previously linearized, or even one may choose to proceed sequentially with the polymerization then with the   /17
ligation of the polymers to the cloning and expression vector.

In both cases, next the transformation or transfection of competent bacterial cells (or cells in culture) is carried out. This step conducts the cloning of the stochastic genes in the living cells where they are indefinitely propagated and expressed.

Of course, beside the operating methods that have just been described, any other appropriate method could also be used for the synthesis of stochastic sequences. In particular, the polymerization could be carried out by the biochemical route with monocatenary oligomers of DNA or RNA obtained by chemical synthesis, then these segments of DNA or RNA treated with methods known in themselves for preparation of copies of bicatenary DNA (c DNA) with a view to cloning the genes.

Screening or selection of the strains of modified host cells

The later step of the method according to the invention consists of examining the transformed or transfected cells, by selection or screening, with a view to isolating one or more cells the transformant or transfectant DNA of which leads to the synthesis of a transcription product (RNA) or translation product (protein) with a desired property. These properties may be enzymatic, functional or structural for example.

One of the most remarkable characteristics of the method according to the invention is enabling simultaneous screening or selection of a usable product (RNA or protein) and of its producing gene. What is more, the DNA synthesized and cloned as described may be selected or screened with a view to isolating the sequences of DNA forming a product in itself, endowed /18 with usable biochemical properties.

Preferential operating methods will now be described by way of nonlimiting examples for screening or selection of strains of transformed cells as well as new proteins presenting a value in view of industrial or medical applications.

One of the operating methods results from the idea of producing a series of polyclonal or monoclonal antibodies obtained in a way known in itself, directed against a protein or another type of molecule of biochemical or medical value, this molecule being, or can be made, immunogenic and to use these antibodies as probes to identify the clones that react with these antibodies among the very numerous clones transformed by stochastic genes. This reaction results from the structure homology existing between the polypeptide synthesized by the stochastic gene and the initial molecule. Thus, numerous new proteins may be isolated that are included as epitopes or antigenic determinants of the initial molecule. Such new proteins are capable of simulating, modulating or blocking the effect of the initial molecule. It will be understood that this method of selection or screening is, in itself, capable of having very numerous pharmacological and biomedical applications. This first operating method will now be described in relation to a specific case and by way of nonlimiting example:

The EGF (epidermal growth factor) is a small protein present in the blood the role of which is to stimulate the growth of the epithelial cells. This effect is obtained by the interaction of the EGF with a specific receptor located in the membrane of the epithelial cells.

Antibodies against EGF are prepared by injection of EGF coupled to KLH (keyhole /19 limpet hemocyanin) to increase the immunogenicity of the EGF. The anti-EGF antibodies of the immunized animals are purified, for example by passage through an affinity column, in which the ligand is EGF or a synthetic peptide corresponding to a fragment of EGF. The purified anti-EGF antibodies are used as probe for screening a large number of bacterial clones lysed with chloroform on solid support. The anti-EGF antibodies are combined with stochastic peptides or proteins, the epitopes of which resemble those of the initial antigen. The clones containing these peptides or proteins are demonstrated by autoradiography after incubation of the solid supports with the radioactive A protein or after incubation with a radioactive anti-antibody.

These steps identify the clones each containing a protein (and its gene) reacting with the screening antibody. Thus screenings can be made among a very large number of bacterial cell strains or viral plaques (for example on the order of a million) and it is possible to detect extremely low quantities, for example on the order of 1 ng, of protein produced. Next, the culturing of the identified clones is carried out, then the purification of the proteins detected by conventional means. The purified proteins are tested in vitro in cultures of epithelial cells to determine if they inhibit, simulate or modulate the effect on these cultures of EGF. Some of the proteins obtained by this means are capable of being used for the chemotherapy treatment of epitheliomas. The activities of the proteins so obtained may be improved by mutation of the DNA coding for the proteins, in a manner similar to that described above. A variant of this operating method consists of purifying stochastic peptides, polypeptides or proteins that may be used as vaccines or more generally may be used for conferring immunity against a pathogenic /20 agent or for exerting other effects on the immunological system, for example, creating a tolerance or decreasing the hypersensitivity in regard to a given antigen, especially as a result of a combination of these peptides, polypeptides or proteins with the antibodies directed against this antigen. It will be understood therefore that these peptides, polypeptides or proteins can be used in vitro as well as in vivo.

More precisely, for example, in the assembly of new proteins which react with the antibody against a given antigen X, each has at least one epitope in common with X, therefore the assembly has an assembly of epitopes in common with X. This makes it possible to use the assembly or a subassembly as a vaccine to confer immunity against X. For example, it is easy to purify one or more of the capsid proteins of the hepatitis B virus. These proteins are injected into the animal, for example the rabbit, and the antibodies corresponding to the starting antigen collected by purification on affinity column. These antibodies are used in the manner described

above to identify the clones producing a protein that has an epitope similar to at least one of the epitopes of the initial antigen. After purification, these proteins are used as an antigen (either alone or in combination) with the purpose of conferring protection against hepatitis B. The later production of the vaccine no longer requires resorting to the initial pathogenic agent.

Note that during the description of the operating methods that precede, a certain number of ways of proceeding with the selection and screening have been described. All these operating methods involve the purification of a specific protein from a transformed strain. These purifications of proteins may be carried out in conformance with the usual methods and in particular, resort to gel chromatography techniques, on ion exchanger and to affinity chromatography. Moreover, the proteins derived from stochastic genes may have been cloned in $/21$ the form of hybrid proteins including for example, a sequence of the β-galactosidase enzyme enabling affinity chromatography on anti-β-galactosidase antibody and enabling the subsequent cleavage of the hybrid part (that is, enabling the new part to be separated from the bacterial part). Next the principle and the implementation of the selection of peptides or polypeptides and corresponding genes expressing these peptides and polypeptides will be described in conformance with a second method of screening or selection based on the detection of the ability of these peptides or polypeptides to catalyze a specific reaction.

By way of nonlimiting practical example, the use of the screening or selection in the particular case of proteins capable of catalyzing the cleavage of lactose, which is normally a function fulfilled by the enzyme β-galactosidase (β-gal).

As described above, the first step in the method consists of creating a large number of expression vectors each expressing a distinct new protein. Practically, for example, the expression vector pUC8 may be chosen with cloning of the stochastic sequences of DNA at the restriction site Pst 1.

The plasmids so obtained are then introduced into a strain of E. coli in the genome of which the natural gene for β-galactosidase, Z, is introduced by conventional genetic methods and also a second gene EBG, without any relationship to the first, but capable of mutating towards the β-gal function. Such host cells (Z⁻, EBG⁻) are not capable of catalyzing the hydrolysis of lactose by themselves and, as a result, of using the lactose as a source of carbon for growth. This makes it possible to use this host strain for screening or selection of the β-gal function.

A biological test method that is suitable for the study of transformed strains of E. coli $/22$ with new genes expressing the β-gal function consists of the culture of bacteria so transformed in Petri dishes containing the X-gal medium. In this case, every bacterial colony expressing a β-gal function is visualized in the form of a blue colony. By using such a biological test, even a weak catalytic activity can be detected. The specific activity of the most characteristic enzymes is established between 10 and 10,000 molecules of product per second.

By assuming that a protein synthesized by a stochastic gene presents a low specific activity, on the order of one molecule per 100 sec, it would still be possible to detect such a catalytic activity. In a Petri dish containing the X-gal medium, in the presence of the non-metabolizable inducer IPTG (isopropyl-D-thiogalactoside) the visualization of a blue region requires the cleavage of approximately 10^{10}-10^{11} molecules of X-gal per square millimeter. A bacterial colony expressing a weak enzyme and occupying a surface area of 1mm^2 contains approximately 10^7-10^8 cells. If each cell presents a single copy of the weak enzyme, each cell should catalyze between 10,000 and 100 cleavages of X-gal in order to be able to be detected which takes approximately 2.7-270 h. Given that, under selective conditions, one may expect an amplification of the number of copies of plasmids per cells for example from 5-20 copies per cell and even from 100-1000 and because of the fact that up to 10% of the proteins of the cell may be specified by the new gene, the length of time necessary for the detection of a blue colony in the case of 100 molecules of enzymes with low specific activity per cell would be on the order of 0.27 h to 2.7 h.

As a result, the screening of a large number of independent bacterial colonies each expressing a different new gene, by taking as selection criterion the ease of expressing the β-gal function is perfectly achievable. Thus, the screening of approximately 2000 colonies in an ordinary Petri dish that is 10 cm in diameter can be carried out. As a result, approximately 20 million colonies can be screened on a 1-m^2 sheet of X-gal agar.

Note that the bacterial colonies appearing blue on the Petri dishes could be falsely positive because of a mutation in the bacterial genome conferring on it the ability to metabolize lactose or for reasons other than those that result from the catalytic activity of the new protein expressed by the cells of the colony. Such false positives may be directly removed by purifying the DNA of the expression vectors originating from the positive colonies and retransforming the E. coli Z$^-$, EBG$^-$ host cells with it. If the β-gal activity results from the new protein coded by the new gene in the expression vector, all the cells transformed by this vector would present the β-gal function. On the other hand, if the initial blue colony results from a mutation in the genome of the host cell, it is an exceptional event independent of the transformation and the number of cells of the new transformed E. coli strain acting as host capable of expressing the β-gal function would have to be low or none.

Specifically, the power of simultaneous purification of expression vectors in bulk for all the positive clones (blue) followed by the retransformation of naïve bacteria will be emphasized. Assume that one wants to carry out screening with a view to selecting proteins with a catalytic function and that the probability that a new peptide or polypeptide fulfills this function, at least weakly, is 10^{-6}, while the probability that the E. coli host microbial strain is subject to a mutation which has the effect of becoming capable of fulfilling this function, is 10^{-5}, it may be calculated

that out of 20 million bacteria transformed that are subjected to screening, on average, 20 positive clones will be attributable to new genes on the expression vectors that each includes while 200 positive clones will result from background mutations. The purification in bulk of the expression vectors with 220 in all and the retransformation of the naïve bacteria with expression vectors pooled will produce a large number of positive clones formed from all the bacteria transformed with the 20 expression vectors that code for the new proteins with the desired function, and a very small number of bacterial clones resulting from background mutations containing the 200 expression vectors remaining without value. A small number of purification cycles of expression vectors in the positive bacterial colonies, as well as from retransformation, enables the detection of very rare truly positive expression vectors compared to a desired catalytic activity, in spite of a very high background of mutations of the host cells for this function.

Following screening operations of this type, the new protein may be purified as a result of the use of usual techniques. The production of this protein in large quantity is made possible as a result of the fact that the identification of the useful protein is accompanied by the simultaneous identification of the gene coding for this protein. As a result, the expression vector itself may be used or even the new gene may be transplanted into an expression vector more appropriate for the synthesis and isolation in large quantity.

Methods of screening of this type may be applied to any enzyme function for which there is an appropriate biological test. Such screenings are not required when the enzymatic function that is sought is profitable to the host cell. Screening may be carried out not only compared with an enzymatic function but also for any other desired property for which an appropriate biological test can be established. Thus, even in the simplest case of the β-gal function visualized by in the Petri dish with X-gal medium, the screening of a number of new genes on the order of 100 million or one billion may be carried out for catalytic activity or another desired property.

/25

Selection of modified host cells

On the other hand, selection techniques for any property, catalytic or otherwise, may be used, the presence or absence of which may be made essential to the survival of the host cells containing the expression vectors coding for the new genes or being able to be useful for the selection of viruses coding and expressing the new gene. By way of nonlimiting practical example, the description of the selection method in relation to the β-galactosidase function will now be reviewed. An appropriate Z⁻EBG⁻ strain of E. coli cannot grow by using lactose as its only source of carbon. Thus, after implementing the first step described above, a very large number of transformed hosts may be cultured by expression vector coding the new genes, the culture being carried out under selective conditions, either by progressively decreasing the

concentration of other carbon sources or by using only lactose from the start. During such a selection, the mutagenesis in vivo by the recombination route or by explicit recovery of expression vectors and mutagenesis in vitro of their new genes by varied mutagens or by another usual technique enables adaptive improvements in the ability to fulfill the desired catalytic function. When there are selection techniques and easy to use techniques at the same time as in the case of the present example, initially selection techniques may be used to enrich the representation in host bacteria expressing the β-gal function, then carrying out screening in a Petri dish on X-gal medium in order to effectively identify positive cells. In the absence of a suitable biological technique, the application of more and more rigorous selection conditions forms the easiest route to purify a type or a small number of types of distinct host cells the expression vectors of which code for proteins catalyzing the chosen reaction.

/26

These techniques may be used to find new proteins with a large variety of structural or functional characteristics in addition to the ability to catalyze specific reactions. For example, screening or selection may be carried out to find new proteins being attached to cis-regulator sites of the DNA and because of this blocking the expression of a function of the host cells, or even blocking the transcription of DNA, stimulating transcription etc.

For example, in the case of E. coli a mutant strain of the repressor of the lactose operon (i⁻) expresses the β-gal function constitutively because of the fact that the lactose operator is not repressed. All the cells of this type produce blue clones on the Petri dishes containing the X-gal medium. It is possible to transform such host strains with expression vectors synthesizing new proteins and to conduct screening in a Petri dish on X-gal medium with a view to detecting clones which are not blue. Among the latter, some represent cases where the new protein is fixed on the lactose operator and represses the synthesis of β-gal. Such plasmids may be purified in bulk, retransformed, clones isolated that do not produce β-gal, and then a detailed verification carried out.

As mentioned above, the method may be used for purposes of creating then isolating not only usable proteins but also RNAs and DNAs forming products in themselves that are provided with usable properties. This obviously results from the fact that on the one hand, the method consists of creating stochastic sequences of DNA that are capable of interacting directly with other cell or biochemical components and that on the other hand, these sequences cloned in an expression vector are transcribed in RNA that are also capable of multiple biochemical interactions.

27

Example of use of the method for the creation and selection of a DNA usable in itself:

This example illustrates the selection of a usable DNA and the purification and study of the mode of action of regulatory proteins being bound to the DNA.

That is to say a preparation of the estradiol receptor, a protein obtained by a method known in itself. In the presence of estradiol, a steroid sex hormone, this receptor changes conformation and is strongly bound to some specific sequences of the genome DNA, thus affecting the transcription of genes involved in sex differentiation and control of fertility.

By incubating a mixture formed of estradiol, its receptor and a large number of different stochastic sequences inserted in their vector then by filtering the mixture through a nitrocellulose membrane, a direct selection of stochastic sequences is obtained that are bound to the estrogen-receptor complex since only the DNAs bound to a protein are retained by the membrane. After washing and elution, the DNA released from the membrane is used as such to transform bacteria. After culture of the transformed bacteria, the vectors that they contain are purified again and one or more cycles of incubation, filtration, transformation are carried out, as described above. These operations make it possible to isolate stochastic sequences of DNA provided with a high affinity for the estradiol-receptor complex. Such sequences are capable of /28 numerous diagnostic and pharmacological applications, in particular for the development of synthetic estrogens for the control of fertility and treatment of sterility.

Creation and selection of a DNA usable in itself

That is to say a large number of sequences of stochastic DNA produced as has been described and cloned in an expression vector. It goes without saying that the RNA transcribed from these sequences in the transformed host cells may be a product usable in itself.

By way of nonlimiting example, a stochastic gene may be selected that codes for a suppressor transfer RNA (t-RNA) by the following procedure:

A competent bacterial strain that includes a "nonsense" mutation in the arg E gene is transformed by a large number ($\geq 10^8$) of stochastic sequences. The transformed bacteria are spread on minimal medium without arginine and containing the selection antibiotic for the plasmid (ampicillin if it the pUC8 vector). Only the transformed bacteria that have become capable of synthesizing arginine could grow. This phenotype may result either from a reverse mutation or form the introduction of a suppressor into the cell. It is easy to test each transformed colony to determine whether or not the arg^+ phenotype results from the presence of the stochastic gene in its vector: one merely prepares the plasmid of this colony and verifies that it confers the Arg^+ phenotype to any arg E cell that it transforms.

Selection of proteins capable of catalyzing a reaction sequence

Another selection method will now be described which is capable of independent applications, based on the principle of simultaneous selection, in parallel, of a certain number of /29 new proteins capable of catalyzing a sequence of reactions connected to each other.

The parent idea of this method is the following: given an assembly of starting chemical compounds considered to be "bricks" or building elements from which one wants to carry out the synthesis of one or more desired chemical compounds through a sequence of catalyzed chemical reactions, there is a very large number of reaction pathways that could be completely or partly substituted for each other, which are all possible from the thermodynamic point of view and which lead to the assembly of "bricks" or building blocks for the desired target chemical compound. Effective synthesis of a target compound is promoted if each step of at least one of the reaction paths leading to the assembly of "building blocks" for the target compound is formed of reactions that are each catalyzed. However, it is relatively less important to determine among the numerous completely or partially reaction pathways those that benefit from catalysis. In the preceding description, it was shown how it was possible to obtain a very large number of host cells which each express a distinct new protein.

Each of these new proteins is capable of catalyzing any one of the possible reactions in the assembly of all the possible reactions leading from the assembly of building blocks to the target compound. If a sufficiently large number of stochastic proteins is present in a reaction mixture containing the compounds forming the building blocks, such that a sufficiently high number of possible reactions is catalyzed, there is a strong probability that a sequence of reaction connected to each other to lead to the assembly of building blocks for the target compound will be catalyzed by a subassembly of new proteins. It is obvious that the method may be extended to the catalysis not only of one but several target compounds simultaneously.   /30

By being based on the principle that has just been explained, one can proceed in the following way to select an assembly of new proteins in parallel that catalyze a desired sequence of chemical reactions:

1. Specify the desired group of compounds forming the "building blocks" by using, preferably, a reasonably high number of distinct chemical species to increase the number of concurrent potential pathways leading to the desired target chemical compound.

2. In an appropriate volume of a solution of reaction medium, add a very large number of new stochastic proteins isolated from transformed or transfected cells synthesizing these proteins. Carry out a test to determine if the target compound is formed. If it is, confirm that this formation requires the presence of the mixture of new proteins. If this is the case, this mixture must contain a subassembly of proteins catalyzing one or more reaction sequences leading from the assembly of "building blocks" to the target compound. Purify by dividing the initial reserve of clones which synthesize the assembly of new proteins. [sic] stochastic, the subassembly necessary for catalyzing the sequence of reactions leading to the target product.

More specifically, by way of nonlimiting example, the selection of an assembly of new proteins will be described that are likely to catalyze the synthesis of a small specified peptide,

namely a pentapeptide, from an assembly of building blocks formed from smaller peptides and amino acids. Every peptide is formed by a linear sequence of 20 different types of amino acids, oriented from its amino end to its carboxy end. Every peptide may be formed in one step by terminal condensation of two smaller peptides (or two amino acids) or by hydrolysis of a larger peptide. A peptide comprising M residues could therefore be formed from an equal number of M-1 condensation reactions. The number of reactions, R, by which an assembly of peptides having a length of 1, 2, 3....M residues, may be interconverted is larger than the number of · molecular species T. This is expressed R/T ≈M-2. Thus by starting from a given group of peptides, a large number of independent or partially independent reaction paths leads to the synthesis of a specific target pentapeptide. A pentapeptide may be chosen in which the presence may be easily detected as a result of a test carried out by usual techniques, for example by HPLC analysis (high pressure liquid chromatography), paper chromatography, etc. The formation of the peptide bond requires energy in dilute aqueous medium but, if the peptides participating in the condensation reactions are sufficiently concentrated, it is the formation of the peptide bond rather than the hydrolysis that is found to be thermodynamically favored and which is produced with a high yield in the presence of an appropriate enzyme catalyst, for example, pepsin or trypsin, without requiring the presence of ATP or other high energy compounds. Such a reaction mixture of small-size peptides may be used and in which the amino acids are radioactively labeled, by means of radioactive tracers of the ^3H, ^{14}C, ^{35}S type to form the assembly of building blocks at sufficiently high concentration to lead to condensation reactions.

/31

For example, one may proceed in the following way:

approximately 15 mg of each amino acid and a small peptide with 2-4 amino acids that are chosen to form the assembly of building blocks are dissolved in a volume of 0.25 mL to 1.0 mL of 0.1M phosphate buffer, pH 7.6. A large number of new proteins created and isolated as described above from their bacterial host or otherwise are purified. The mixture of these new proteins is dissolved up to a final total concentration on the order of 0.8-1.0 mg/mL in the same buffer. 0.25 mL to 0.5 mL of the mixture of proteins is added to the mixture of building blocks. It is incubated at a temperature of 25 °C to 40 °C for 1-40 h. Aliquots of 8 µL are removed at regular intervals, the first of which is considered to be a "blank control" before addition of the mixture of new proteins. The samples are subjected to a chromatography analysis by using as solvent a mixture of n-butanol-acetic acid-pyridine-water (30:6:20:24 by volume). The chromatogram is dried and revealed with ninhydrin or autoradiography (with or without intensification screen). Because the compounds forming the building blocks are radioactively labeled, the target compound will be radioactive and it will have a high specific activity that enables its detection at 1-10 ng. Instead of the usual chromatographic analysis, a HPLC analysis (high pressure liquid chromatography) may be carried out; it is quicker and simpler to carry out.

/32

Generally, the usual analysis methods may be used. In that way, a yield of target compounds may be detected that is less than one part per million by weight compared with the initial or "building material" compounds.

In the case where the pentapeptide is formed under the conditions described above but where its formation has not taken place when an extract purified as above is used, but obtained from cells transformed by the expression vector lacking stochastic insertion, the formation of the pentapeptide does not result from the presence of the bacterial contaminants and therefore requires the presence of a subassembly of new proteins in the reaction mixture.

The following step consists of the separation of the particular subassembly of cells that /33 contain the expression vectors of new proteins catalyzing the sequence of reactions leading to the target pentapeptide. For example, if the number of reactions forming this sequence is equal to 5, there are approximately 5 new proteins that catalyze the necessary reactions. In the case where the "bank of clones" of bacteria containing the expression vectors coding for the new genes contain a number of distinct new genes on the order of 1,000,000, the isolation is carried out in bulk of all these expression vectors and the retransformation of 100 distinct assemblies of 10^8 bacteria with a sufficiently low ratio of vectors to bacteria of that on average each group of bacteria are transformed by only approximately half the initial number of genes, that is approximately 500,000. As a result, the probability that any of the 100 assemblies of bacteria contain all 5 new critical proteins is equal to $(1/2)^5 = 1/32$. Among the 100 initial assemblies of bacteria, approximately 3 will contain the 5 critical transformants. In each of these assemblies, the total quantity of new genes present is no more than 500,000 instead of one million. By successive repetitions of this procedure, the total number of which is 20 in the present case the 5 critical new genes can be isolated. After which, the mutagenesis and the selection of this group of 5 stochastic genes makes it possible to search for improved catalytic functions. In the case where it is necessary to catalyze a sequence of reactions comprising a number of reactions on the order of 20 and when 20 genes coding for new proteins must be isolated in parallel, it is sufficient to adjust the multiplicity of transformations such that each group of 10^8 bacteria receives 80% of the 10^6 stochastic genes by using 200 assemblies of this type. The probability that the 20 new proteins are found in a given group of bacteria is 0.8^{20}, that is, approximately 0.015. As a result, 2 of the 200 assemblies will contain the 20 new genes necessary to catalyze the formation of the target compound. The number of repetitions of cycles necessary for the /34 isolation of the 20 new genes is on the order of 30.

The principles and the operating methods mentioned above are widespread in the case of peptides in numerous fields of chemistry in which reactions in aqueous media may be carried out under conditions of pH, temperature and concentrations of solutions that allow the general enzymatic functions. In each case, it is necessary to be able to provide a test method to detect the

formation of the desired target compound(s). it is also necessary to choose a sufficiently high number of "building blocks" to increase the number of reaction sequences that lead to the target compound.

The practical example that has just been given of the synthesis of a target pentapeptide may be brought to general use in the following way:

The process as it is described creates stochastic peptides and proteins among other products. These peptides or proteins may act by the catalytic route or otherwise on other compounds. They may also form the substrates on which they act. Also, one may select (or screen) the ability of the stochastic peptides or proteins to interact among themselves and because of this to modify the conformation, structure or function of some of them. Likewise, the ability of these peptides or proteins to catalyze among themselves the hydrolysis, condensation, transpeptidation or other modification reactions may be selected (or screened). For example, the hydrolysis of a given stochastic protein by at least one member of the assembly of stochastic peptides and proteins may be followed and measured by radioactive labeling of the given protein followed by incubation with the mixture of stochastic proteins, in the presence of such ions as Mg, Ca, Zn, Fe and compounds ATP and GTP. Next, the appearance of radioactive fragments of   /35 the labeled protein is measured as has been described. The stochastic protein(s) that catalyze this reaction may then be isolated, as well as their producing gene by sequential decrease of the library of transformant clones as described.

An extension of the method consists of selecting an assembly of stochastic peptides and polypeptides that are capable of catalyzing a continuation of the reactions leading from the starting components (amino acids and small peptides) to some peptides or polypeptides of the assembly. Thus, it is also conceivable to select an assembly capable of catalyzing its own synthesis: such a reflexively autocatalytic group may be set up in a chemstat where the reaction products are constantly diluted but where the concentration of starting products is kept constant. The existence of an assembly of this type may be verified by two dimensional gel chromatography and by "HPLC" showing the synthesis of a stable distribution of peptides and polypeptides. The reaction volumes depend on the number of molecular species used and the concentrations necessary for promoting the formation of peptide bonds compared to the hydrolysis. The distribution of molecular species of an autocatalytic group is capable of varying or deriving subsequent to the emergence of varying autocatalytic assemblies. The peptides and polypeptides that form an autocatalytic group may have some parts in common with the vast starting group (formed of peptides and polypeptides coded according to the method), but may also contain peptides and polypeptides not coded by the assembly of stochastic genes coding the starting group.

All the stochastic genes for which the products are necessary to establish such an autocatalytic group may be isolated as has been described, by sequential decreases in the library of transformant clones. Moreover, an autocatalytic group may contain peptides initially coded by /36 the stochastic genes and formed continuously in the autocatalytic group. To isolate this coded subassembly of peptides and polypeptides the autocatalytic group may be used to obtain polyclonal sera by immunization in animals that recognize a very wide number of components of the autocatalytic group.

These sera may then be used to screen the library of stochastic genes to find the genes in it that express proteins capable of being combined with antibodies present in the sera.

This group of stochastic genes expresses a large number of coded stochastic proteins that persist in the autocatalytic group. The rest of the coded components of such an autocatalytic group may be isolated by sequential decrease, as described, of the library of stochastic genes from which the subassembly detected by the immunological method has been removed.

The autocatalytic assemblies of peptides and proteins obtained in the manner described above are likely to find numerous practical applications.

Claims   /37

1. Production method of peptides or polypeptides by the microbiological route, characterized in that simultaneously genes are produced within the same medium that are at least partially composed of stochastic synthetic polynucleotides, that the genes so obtained are introduced into host cells, that the independent strains of modified host cells containing these genes are cultured simultaneously so as to clone the stochastic genes and to promote the production of genes expressed by each of the stochastic genes, that the screening and/or selection of the strains of modified host cells is carried out so as to identify the strains producing peptides or polypeptides with at least one given property, that the strains so identified are isolated and that they are cultured so as to produce at least one peptide or polypeptide with said property.

2. Method according to Claim 1 characterized by the fact that the genes are produced by stochastic copolymerization from four types of deoxyphosphononucleotides A, C, G and T from two ends of a previously linearized expression vector, then formation of cohesive ends so as to form a first strand of stochastic DNA formed from a molecule of the expression vector containing two stochastic sequences the 3' ends of which are complementary, followed by the synthesis of the second strand of this stochastic DNA.

3. Method according to Claim 1 characterized by the fact the genes are produced by stochastic copolymerization of bicaternary oligonucleotides without cohesive ends, so as to form fragments of stochastic DNA, followed by ligation of these fragments to a previously linearized expression vector.

4. Method according to Claim 2 or Claim 3 characterized by the fact that the expression /38
vector is a plasmid.

5. Method according to Claim 4 characterized by the fact the expression vector is the
plasmid pUC8.

6. Method according to Claim 2 or Claim 3 characterized by the fact that the expression
vector is a fragment of viral DNA.

7. Method according to Claim 2 or Claim 3 characterized by the fact that the expression
is a hybrid of plasmid and viral DNA.

8. Method according to one of Claims 1-6 characterized by the fact that the host cells are
prokaryotic cells.

9. Method according to one of Claims 1-7 characterized by the fact that the host cells are
eukaryotic cells.

10. Method according to Claim 8 characterized by the fact the cells are chosen from HB
101 and C 600.

11. Method according to Claim 3 characterized by the fact that the oligonucleotides form
an assembly of palindromic octamers.

12. Method according to Claim 11 characterized by the fact that the assembly of
palindromic octamers is the following group:

```
5'  GGAATTCC  3'
5'  GGTCGACC  3'
5'  CAAGCTTG  3'
5'  CCATATGG  3'
5'  CATCGATG  3'
```

13. Method according to Claim 3 characterized by the fact the oligonucleotides form an /39
assembly of palindromic heptamers.

14. Method according to Claim 13 characterized by the fact that the assembly of
palindromic heptamers is the following group:

```
5'  XTCGCGA  3'
5'  XCTGCAG  3'
5'  RGGTACC  3'
```

where x = A, G, C or T and R = A or T.

15. Method according to Claim 4 and one of Claims 12-14 characterized by the fact the transformant DNA of the plasmids originating from a culture of independent strains of modified host cells obtained by proceeding in the manner specified in Claim 11 or Claim 13 is isolated and purified, then the cleavage of the DNA is promoted by means of at least one restriction enzyme corresponding to a specific enzymatic cleavage site present in these palindromic octamers or heptamers but absent from the expression vector used, this cleavage being followed by the inactivation of the restriction enzyme, and then all the fragments of linearized stochastic DNA so obtained by the T4 DNA ligase are treated simultaneously so as to create a new DNA assembly that contains new stochastic sequences, and this new assembly of transformant DNA is used to modify host cells and clone genes, and finally, the new strains of transformed host cells are screened and/or selected and isolated and finally they are cultured so as to produce at least one peptide or polypeptide.

16. Method according to Claim 1 characterized by the fact that said property is the ability to catalyze a given chemical reaction.

17. Method according to Claim 1 for the production of several peptides and/or polypeptides characterized by the fact that said property is the ability to catalyze a sequence of reactions leading from a given initial group of chemical compounds to at least one target compound. /40

18. Method according to Claim 1 for the production of an assembly formed from several peptides and/or polypeptides reflexively autocatalytic, characterized by the fact that said property is the ability to catalyze the synthesis of this assembly itself from amino acids and/or oligopeptides.

19. Method according to Claim 1 characterized by the fact said property is the ability to selectively modify the chemical and/or biological properties of a given compound.

20. Method according to Claim 19 characterized by the fact that said property is the ability to selectively modify the catalytic activity of a polypeptide.

21. Method according to Claim 19 characterized by the fact that said property is the ability to simulate, inhibit, or modify at least one biological function of at least one biologically active compound.

22. Method according to Claim 21 characterized by the fact that said biologically active compound is chosen from hormones, neurotransmitters, adhesion factors or growth and the specific regulators of DNA replication and/or transcription and/or translation of RNA.

23. Method according to Claim 1 characterized by the fact that said property is the ability to be bound to a given ligand.

24. Use of the peptide or polypeptide obtained by the method according to Claim 23 for /41

the detection and/or titration of a ligand.

25. Method according to Claim 1 characterized by the fact that said property is having at least one epitope similar to one of the epitopes of a given antigen.

26. Method according to Claims 19 and 25 characterized by the fact that said property is the ability to simulate or modify the effects of a biologically active molecule, and that the screening and/or selection of the modified strain of host cells producing at least one peptide or polypeptide having this property by preparing antibodies against this active molecule, and by using these antibodies after their purification to identify the strains containing this peptide or polypeptide, then by culturing the strains so identified and by separating and purifying the peptide or polypeptide produced by these strains and finally, by subjecting this peptide or polypeptide to an in vitro test to verify that it really has the ability to simulate or modify the effects of said molecule.

27. Polypeptides obtained by the method according to Claim 1 or Claim 26 usable as an active substance with a pharmacological and/or chemotherapeutic action.

28. Peptides or polypeptides obtained by the method according to Claim 25 usable to decrease, in vitro or in vivo, the concentration of free antibodies, specific against said antigen, by formation of bonds between these peptides or polypeptides and antibodies.

29. Peptides or polypeptides according to Claim 27 or Claim 28 usable as suppressor agent of immune hypersensitivity.

30. Peptides or polypeptides obtained by the method according to Claim 25 usable as agent for creation of a tolerance with regard to said antigen.

31. Method according to Claim 35 characterized by the fact that the antigen is EGF.

32. Polypeptides obtained by the method according to Claim 31 usable for the chemotherapy treatment of epitheliomas.

33. Method according to Claim 1 characterized by the fact that the modified strains of host cells producing the peptides or polypeptides with the desired property are identified and isolated by affinity chromatography on antibodies corresponding to a protein expressed by the natural part of the hybrid DNA.

34. Method according to Claim 33 characterized by the fact that the natural part of the hybrid DNA contains a gene expressing β-galactosidase, said modified strains of host cells may be advantageously identified by affinity chromatography on anti-β-galactosidase antibodies.

35. Method according to Claim 1 or Claim 34 characterized by the fact that after expression and purification of the hybrid peptides or polypeptides their new parts may be separated and isolated.

36. Application of the method according to Claim 25 or Claim 26 for the preparation of a vaccine characterized by the fact that antibodies against a pathogenic agent are isolated, and they

are used to identify the clones producing at least one protein having at least one epitope similar to one of the epitopes of the pathogenic agent, the strains of modified host cells corresponding to these clones are cultured to produce this protein, this protein is isolated and purified from cultures of these strains of cells, and this protein is used for the production of a vaccine against the pathogenic agent.

37. Application according to Claim 36 for the preparation of an anti-HBV vaccine, characterized by the fact that an HBV virus capsid protein can be extracted and purified and this protein injected into the body of an animal capable of forming antibodies against this protein, the antibodies so formed are collected and purified, these antibodies are used to identify the clones producing at least one protein with at least one epitope similar to one of the epitopes of the HBV virus, the strains of modified host cells corresponding to these clones are cultured, so as to produce this protein, this protein is isolated and purified from cultures of these strains of cells, and this protein used for production of an anti-HBV vaccine.

38. Method according to Claim 1 characterized by the fact that the host cells consist of bacteria of the genus Escherichia coli, the genome of which contains neither the natural gene expressing β-galactosidase nor the EBG gene, that is, (Z⁻, EBG⁻) E. coli bacteria, the modified host cells are cultured in the presence of the X-gal medium and the IPTG inducer, that the clones positive for the β-galactosidase function are detected in the culture medium and finally, then this DNA is transplanted into an appropriate strain of host cells for culture in large quantity with a view to the industrial production of at least one peptide, polypeptide or protein.

39. Method according to Claim 1 characterized by the fact that said property is the ability of the polypeptides or peptides to be bound to a given compound.

40. Method according to Claim 39 characterized by the fact that said compound is chosen from the peptides, polypeptides and proteins.

41. Method according to Claim 40 characterized by the fact that said proteins are regulatory proteins of DNA transcription or replication activity.

42. Method according to Claim 39 characterized by the fact said compound is chosen from the DNA and RNA sequences.

43. Proteins obtained by the method according to Claim 40 or Claim 42.

44. Method for production of DNA characterized by the fact that simultaneously and within the same medium genes are produced that are at least partially composed of stochastic synthetic polynucleotides, that the genes so obtained are introduced into host cells, so as to produce an assembly of modified host cells, that simultaneously the independent strains of modified host cells so produced are cultured, that screening and/or selection of this group is carried out in order to identify the host cells that contain desired sequences in their genome of

DNA stochastic sequences presenting at least one desired property and that the DNA is isolated from cultures of host cells so identified.

45. Method according to Claim 44 characterized by the fact that said property is the ability to be bound to a given compound.

46. Method according to Claim 45 characterized by the fact said compound is chosen from peptides, polypeptides and proteins.

47. Method according to Claim 45 characterized by the fact said compound is a regulator compound of the DNA transcription or replication activity.

48. Method according to Claim 47 characterized by the fact said compound is a regulatory protein of DNA transcription or replication activity. /45

49. Use of a sequence of DNA obtained by the method according to Claim 46 or Claim 47 as cis-regulatory sequence of the replication or transcription of a close DNA sequence.

50. Method according to Claim 42 characterized by the fact the proteins obtained present in addition the property of modifying the activity of DNA transcription or replication or stability.

51. Use of a protein obtained by the method according to Claim 48 to modify the transcription or replication or stability properties of a DNA sequence in a cell containing this sequence of DNA and expressing this protein.

52. Production method of RNA characterized by the fact that simultaneously and within the same medium, genes are produced that are at least partially composed of stochastic synthetic polynucleotides, that the genes so obtained are introduced into host cells, so as to produce an assembly of modified host cells, that the independent strains of modified host cells so produced are simultaneously cultured, that screening and/or selection of this group is carried out in order to identify the host cells that contain stochastic sequences of RNA with at least one desired property, and that the RNA is isolated from cultures of host cells so identified.

53. Method according to Claim 52 characterized by the fact said property is the ability to be bound to a given compound.

54. Method according to Claim 52 characterized by the fact that said property is the ability to catalyze a given chemical reaction. /46

55. Method according to Claim 52 characterized by the fact that said property is of being a transfer RNA.

INTERNATIONAL SEARCH REPORT

International Application No   PCT/CH 85/00099

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) °

According to International Patent Classification (IPC) or to both National Classification and IPC

Int. Cl.^{4} : C 12 N 15/00; C 12 P 21/00; C 07 K 15/00; G 01 N 33/48; A 61 K 37/02; A 61 K 39/00; A 61 K 39/29

II. FIELDS SEARCHED

Minimum Documentation Searched ⁷	
Classification System	Classification Symbols
Int. Cl.^{4}	C 12 N

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched ⁸

III. DOCUMENTS CONSIDERED TO BE RELEVANT ⁹

Category *	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
A	DE, A, 3246071 (CHRISTIAN STROBEL) 14 June 1984, see claims 1-16; page 8, figure 7 to page 9; figure 14	1
A	DE, A, 3303173 (CHRISTIAN STROBEL) 02 August 1984, see claims 1-10	1
A	DE, A, 3300632 (CHRISTIAN STROBEL) 12 July 1984, see claims 1-23	1
A	Chemical Abstracts, Vol. 84, No. 5, 02 February 1976, Columbus, Ohio (US) G.W.Hoffmann : 'Stochastic theory of the origin of the genetic code', see page 136, abstract 26960x, & Annu. Rev. Phys. Chem. 1975, 26, 123-44 (Eng.)	1

* Special categories of cited documents: ¹⁰

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
25 November 1985 (25.11.85)	12 December 1985 (12.12.85)
International Searching Authority	Signature of Authorized Officer
European Patent Office	

Form PCT/ISA/210 (second sheet) (January 1985)

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON

INTERNATIONAL APPLICATION NO.   PCT/CH 85/00099 (SA   9773)
--------------------------------   ----------------- ------------

This Annex lists the patent family members relating to the
patent documents cited in the above-mentioned international
search report. The members are as contained in the European
Patent Office EDP file on 03/12/85

The European Patent Office is in no way liable for these
particulars which are merely given for the purpose of
information.

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
DE-A- 3246071	14/06/84	DE-A-	3300632	12/07/84
		DE-A-	3303173	02/08/84
DE-A- 3303173	02/08/84	DE-A-	3246071	14/06/84
		DE-A-	3300632	12/07/84
DE-A- 3300632	12/07/84	DE-A-	3246071	14/06/84
		DE-A-	3303173	02/08/84